

# Exploring Spiking Neural Networks for Binary Classification in Multivariate Time Series at the Edge

James Ghawaly Jr.

*Computer Science and Engineering  
Louisiana State University  
Baton Rouge, United States  
jghawaly@lsu.edu*

Andrew Nicholson

*Physics  
Oak Ridge National Laboratory  
Oak Ridge, United States  
nicholsonad@ornl.gov*

Catherine Schuman

*Electrical Engineering and Computer Science  
The University of Tennessee Knoxville  
Knoxville, United States  
cschuman@utk.edu*

Dalton Diez

*Computer Science and Engineering  
Louisiana State University  
Baton Rouge, United States  
ddiez2@lsu.edu*

Aaron Young

*Computer Science and Mathematics  
Oak Ridge National Laboratory  
Oak Ridge, United States  
youngar@ornl.gov*

Brett Witherspoon

*Electrification and Energy Infrastructures  
Oak Ridge National Laboratory  
Oak Ridge, United States  
witherspoocb@ornl.gov*

**Abstract**—We present a general framework for training spiking neural networks (SNNs) to perform binary classification on multivariate time series, with a focus on step-wise prediction and high precision at low false alarm rates. The approach uses the Evolutionary Optimization of Neuromorphic Systems (EONS) algorithm to evolve sparse, stateful SNNs by jointly optimizing their architectures and parameters. Inputs are encoded into spike trains, and predictions are made by thresholding a single output neuron’s spike counts. We also incorporate simple voting ensemble methods to improve performance and robustness.

To evaluate the framework, we apply it with application-specific optimizations to the task of detecting low signal-to-noise ratio radioactive sources in gamma-ray spectral data. The resulting SNNs, with as few as 49 neurons and 66 synapses, achieve a 51.8% true positive rate (TPR) at a false alarm rate of 1/hr, outperforming PCA (42.7%) and deep learning (49.8%) baselines. A three-model any-vote ensemble increases TPR to 67.1% at the same false alarm rate. Hardware deployment on the  $\mu$ Caspian neuromorphic platform demonstrates 2 mW power consumption and 20.2 ms inference latency.

We also demonstrate generalizability by applying the same framework, without domain-specific modification, to seizure detection in EEG recordings. An ensemble achieves 95% TPR with a 16% false positive rate, comparable to recent deep learning approaches with significant reduction in parameter count.

**Index Terms**—neuromorphic computing, spiking neural networks, anomaly detection

## I. INTRODUCTION

The rapid advancement of generative Deep Learning (DL) technology has significantly transformed human interaction with information, enabling the automation of diverse tasks. However, this progress has highlighted critical limitations,

This research was supported by the U.S. National Nuclear Security Administration (NNSA) Office of Defense Nuclear Nonproliferation Research and Development within the U.S. Department of Energy under Contract AC05-00OR22725.

including the substantial energy requirements of large-scale models [1], reliance on increasingly scarce and expansive datasets [2], and challenges in achieving robust reasoning and adaptability [3]–[5]. Neuromorphic computing, inspired by the proven efficiency and adaptability of biological neural systems, offers a potential pathway to address these challenges [6]–[9]. By leveraging the demonstrated success of biological neural networks as a model for computation, neuromorphic systems aim to replicate their efficiency and versatility. Through hardware-software co-design, these systems integrate energy-efficient computational architectures with biologically plausible Spiking Neural Network (SNN) models, facilitating the development of scalable, sustainable, and adaptive Artificial Intelligence (AI) solutions.

While SNNs currently do not outperform deep learning models in most machine learning applications, they offer significant advantages in resource-constrained scenarios, such as long-term edge deployments, where traditional deep learning approaches may be impractical. SNNs offer the promise of high performance while maintaining low power through event driven computing. This offers significant advantages in resource-constrained scenarios, such as long-term edge deployments, where traditional deep learning approaches may be impractical. A compelling example is binary classification in multivariate time series data, a task critical to many real-world applications such as security monitoring and healthcare diagnostics. Detecting weak signals of interest in time series data requires models capable of identifying subtle, temporally distributed patterns within noisy and high-dimensional signals. At the edge, these challenges are compounded by the need for energy efficiency and real-time processing capabilities, requirements that often exceed the practical limits of conventional deep learning models.

### A. Binary Classification in Multivariate Time Series

We can view a multivariate time series,  $\mathbf{X}$ , as a sequence of temporally ordered observations, represented as

$$\mathbf{X} = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T\},$$

where  $\mathbf{x}_t \in \mathbb{R}^n$  is the observation of  $n$  variables at time  $t$ , and  $T$  is the total number of time steps in  $\mathbf{X}$ . We will discuss two distinct forms of time series binary classification:

*Independent Sample Classification:* The classifier ignores any temporal patterns and assigns a binary label to each observation independently. Each prediction is based solely on the current observation  $\mathbf{x}_t$ . This can be expressed as:

$$f : \mathbf{X} \rightarrow \{0, 1\},$$

where  $f(\mathbf{x}_t)$  predicts the binary label for time  $t$ .

*Stateful Step-Wise Classification:* The classifier assigns a binary label to each time step  $t$ , using only the current observation  $\mathbf{x}_t$  along with an internal state  $\mathcal{H}_t$  that encodes information about prior observations ( $\mathbf{X}_{1:t-1}$ ).  $\mathcal{H}_t$  represents a compressed internal representation of past observations. This allows for more efficient temporal modeling without explicitly storing the entire sequence. This can be formalized as:

$$f : \mathbf{X} \times \mathcal{H} \rightarrow \{0, 1\},$$

where the internal state evolves dynamically over time, being updated recursively based on the prior state and observation. In some cases, such as with Recurrent Neural Networks (RNNs), the state transition is explicitly defined:

$$\mathcal{H}_t = g(\mathcal{H}_{t-1}, \mathbf{x}_{t-1}),$$

where  $g$  is the state transition function. In an SNNs, the state transition function is not explicitly defined but rather emerges from the inherent dynamics of the model, described in more detail in Section III.

### B. Neuromorphic Computing

Neuromorphic systems offer an energy-efficient solution for binary classification in multivariate time-series data. SNNs are naturally suited to this domain due to their stateful dynamics (neurons maintain internal membrane potentials over time), enabling implicit modeling of temporal dependencies at the individual neuron level, without the need to store full input histories (as in transformers) or define special recurrent structures (as in RNNs). Unlike conventional Artificial Neural Networks (ANNs), SNNs encode and process time-varying data through sparse, event-driven spikes. This asynchronous computation also minimizes redundant processing, allowing for low-latency, low-power inference suitable for deployment on edge devices with strict energy and memory constraints [7], [10], [11].

Despite theoretical advantages over traditional ANNs, SNNs have not yet gained widespread adoption in the mainstream machine learning community. One major challenge is determining optimal network architectures for specific tasks, particularly when targeting resource-constrained hardware. Unlike

in ANNs, where well-defined architectural standards exist for many applications, the appropriate design of SNNs is less understood. This challenge is compounded by the need for compact networks that operate efficiently on neuromorphic hardware platforms. While methods like surrogate gradients [12] and EventProp [13] enable the application of gradient descent in specific cases, these approaches require fixed architectures prior to training, which can lead to the loss of some of the potential benefits offered by SNNs.

Evolutionary algorithms, such as Evolutionary Optimization for Neuromorphic Systems (EONS) [14], offer an alternative for developing SNNs. By simultaneously evolving both the architecture and parameters, EONS can explore a wide range of network configurations without requiring predefined structures. Moreover, evolutionary approaches naturally accommodate the event-driven nature of spiking neurons, enabling the creation of compact, efficient networks tailored for deployment on neuromorphic platforms like μCaspian [11].

### C. Focus of this Study

In this work, we leverage EONS to design highly compact SNNs optimized for stateful, step-wise binary classification tasks in multivariate time series. By training these networks directly on hardware simulators within the TENNLab framework, we ensure that the resulting SNNs are not only efficient but also deployable on energy-constrained edge devices.

We propose a general framework and procedure for training and evaluating SNNs for binary classification problems in multivariate time series. Our approach inherently yields SNNs capable of direct deployment on μCaspian [11].

To evaluate the approach, we first conduct an in-depth study on a challenging low-Signal to Noise Ratio (SNR) gamma-ray detection problem. We extend prior work [15], [16] in this domain by introducing improved dataset design and novel training procedures. We then deploy the trained SNN on microCaspian and perform a power analysis, comparing the results against an alternative low-power solution implemented on a Microcontroller Unit (MCU).

Building on the insights gained from gamma-ray detection, we extend our methods to an additional application to assess the generalizability of our framework *without* domain-specific optimization: seizure detection in scalp Electroencephalogram (EEG) recordings.

## II. RELATED WORK

Neuromorphic computing has garnered increasing interest in the research community for its application to anomaly detection in time-series data. One approach that has emerged is the use of Online Evolving Spiking Neural Networks (OESNNs), which identify relationships within input data by capturing temporal patterns from unlabeled normal data and encoding them as the network's normal state. When the input data causes the network's state to deviate from the learned normal state, an anomaly is declared. Studies utilizing OESNNs include the work of [17], which focuses on univariate input data, and [18], which extends the approach to multivariate inputs. These

methods learn by pruning or adding neurons as new data are encountered, so as to maintain optimal performance.

Another proposed method is the *predict and evaluate* approach [19]. In this method, the SNN learns to forecast the next state of the system and then calculates the error between the predicted and actual state. If the error exceeds a predefined threshold, the model declares an anomaly. This approach can be applied to both univariate and multivariate time-series, depending on how the inputs are structured and encoded.

In contrast to these, our work focuses on binary classification using labeled datasets rather than anomaly detection.

The work most related to ours is [20], where the authors propose two SNN models inspired by Reservoir Computing (RC) and Legendre Memory Units, to address univariate time series classification. The first, Spiking Legendre Reservoir Computing (SLRC), follows a conventional RC design and is deployed on Intel’s Loihi platform. The second, Legendre Spiking Neural Network (LSNN), introduces nonlinearity in the readout layer and is trained with Surrogate Gradient Descent (SurrGD). Their models achieve high performance when compared against liquid state machines, however, the authors note that their approach can only be applied to univariate, not multivariate time series. Additionally, our work differs in that we target the  $\mu$ Caspian FPGA-based neuromorphic device, which consumes less power than Loihi, at the expense of reduced maximum network size.

While hardware efficiency is often cited as a motivating factor, true hardware-software co-design is not explored in any of the reviewed prior work, except for that proposed by [20]. In contrast, we train and evaluate our networks directly in a virtual simulator that replicates the  $\mu$ Caspian neuromorphic device. We also deploy the trained SNN on an actual  $\mu$ Caspian device and perform a preliminary power analysis.

Finally, all of the reviewed works focus on optimizing a single SNN for anomaly detection. As discussed in Section I, the design of effective SNN architectures for different applications is poorly understood. Prior work, such as [21], has demonstrated the potential for ensembles of SNNs to improve performance. While they highlight the benefits of ensemble learning, they do not explore methods to evolve individual network architectures. In contrast, our method trains many unique SNNs using EONS and then ensembles up to 3 SNNs.

### III. GENERAL METHODOLOGY

Our proposed methodology develops compact SNNs optimized for binary classification in multivariate time series data, designed for deployment on the low-power  $\mu$ Caspian platform. While application-specific considerations are necessary for optimal performance, the framework is intended to be broadly generalizable as starting point for development. It comprises the following steps: dataset preparation, training on virtualized hardware using EONS within the TENNLab framework, test set evaluation of the trained network population, network ensembling, final model selection and evaluation on hardware. Methods specific to each of the investigated applications are presented in their respective portions of Section IV.

In our framework, each SNN is designed based on the *Stateful Step-wise Binary Classification* paradigm outlined in Section I-A. Specifically, the SNN maps the spike-encoded current observation,  $\mathbf{s}_t$ , and an internal state encoding prior observations,  $\mathcal{H}_{t-1}$ , to a binary class label  $\{0, 1\}$ . Here,  $\mathbf{s}_t$  is a train of spikes of value 1 occurring at set time intervals. Formally, this mapping can be expressed as:

$$f : \mathcal{S} \times \mathcal{H} \rightarrow \{0, 1\},$$

where  $\mathbf{s}_t \in \mathcal{S}$  represents the spike-encoded inputs, and  $\mathcal{H}$  denotes the space of possible internal states.

The internal state at a given time step  $\mathcal{H}_t$  is updated recursively based on the prior state and the previous observation:

$$\mathcal{H}_t = \Psi(\mathcal{H}_{t-1}, \mathbf{s}_{t-1}),$$

where  $\Psi$  represents an abstract function of the internal state dynamics of the SNN. The SNN begins each inference at  $\mathcal{H}_t$  and processes  $\mathbf{s}_t$  through its network of Integrate-and-Fire (IF) neurons for a pre-defined inference time,  $\tau$ . After running for  $\tau$ , the SNN’s state consists of the membrane potentials of its neurons and the spiking activity (in-flight after reaching  $\tau$ ), which depend on the network’s architecture, synaptic weights, neuron thresholds, and previous inputs. Collectively, these processes guide the state transition and encode temporal dependencies, enabling effective processing of sequential data.

For a time series of dimensionality  $n \times T$ , each network has  $n$  or  $2n$  input neurons (one for each variable), depending on the encoder binning scheme selected (see Section III-A), and a single output neuron. The total *spike count* of the output neuron, denoted  $z_t$ , is a non-negative integer representing the number of spikes emitted by the output neuron over the inference period  $\tau$ . We obtain the final binary label  $y_t$  by thresholding:

$$y_t = \begin{cases} 1, & \text{if } z_t > \theta, \\ 0, & \text{otherwise,} \end{cases}$$

where  $\theta$  is a user-defined threshold. Hence, whenever  $z_t$  exceeds  $\theta$ , the SNN predicts a label of 1; otherwise, it predicts 0. In some applications, it may be beneficial to threshold on a rolling sum of spike count with a fixed time window, to threshold on consecutive bursts of spikes. This is beneficial for applications with a high data sampling rate, where spurious noise-induced spikes could increase False Positive Rate (FPR).

#### A. Spike Encoding

For an observation at a given time step,  $\mathbf{x}_t \in \mathbb{R}^n$ , the encoder maps the observation to a corresponding spike train  $\mathbf{s}_t$ . Each component  $x_t^{(i)}$  in  $\mathbf{x}_t$  has a minimum ( $\text{min}_i$ ) and maximum ( $\text{max}_i$ ) value, either derived from the physics of the variable itself or determined empirically from a training dataset that adequately covers the domain of observations. The spike-encoded signal  $\mathbf{s}_t$  thus contains a series of binary pulses over an encoding window of length  $\tau$ , which is the inference time for the SNN.

In prior work on the application discussed in Section IV-A, a hyperparameter exploration procedure was performed across the variety of encoding schemes offered through the TENNLab framework, with two methods demonstrating the highest success: *rate* and *spikes*. We normalize each variable,  $x_t^{(i)}$ , to

$$\tilde{x}_t^{(i)} = \frac{x_t^{(i)} - \min_i}{\max_i - \min_i}.$$

- *Rate Encoding*: Given the encoding window of length  $\tau$ , spikes of value 1 are distributed at a frequency proportional to  $\tilde{x}_t^{(i)}$ . Formally, we model this frequency as  $\tilde{x}_t^{(i)} \times \lambda$ , where  $\lambda$  is the scaling factor indicating the maximum firing rate. In our setting,  $\lambda = \tau$ , meaning that a fully normalized input  $\tilde{x}_t^{(i)} = 1$  corresponds to one spike per unit time, yielding  $\tau$  spikes over the entire inference window. Values above  $\max_i$  saturate to  $\lambda$ .
- *Spikes Encoding*: In this approach, rather than distributing spikes continuously over the entire inference window, the encoder emits spikes at uniform unit-time intervals. The input amplitude is translated into the total number of spikes generated, ranging from 0 up to  $\tau$ .

In our implementation of *spikes* encoding, we incorporate an encoder binning strategy, which distributes spikes into  $b$  neurons (or “bins”), where each bin  $b_i$  handles a specific fractional subrange of  $\min_i$  to  $\max_i$ . This binning process can help preserve finer distinctions in the encoded input while retaining a discrete spiking representation. Finally, the binning strategy leverages a feature called “flip-flop,” wherein even-numbered bins reverse the order of spikes. In other words, for these bins, the maximum firing rate occurs at the minimum value of the corresponding subrange. This design choice can be beneficial for waveforms that oscillate around zero, providing a more balanced representation of positive and negative values.

### B. $\mu$ Caspian

This work leverages the Caspian [11] simulator and the  $\mu$ Caspian [22] Field Programmable Gate Array (FPGA)-based Neuromorphic Event Processor (NEP) both developed at Oak Ridge National Laboratory (ORNL) and designed to integrate as a processor within the TENNLab Framework. The Caspian simulator is a parameterizable event-based neuromorphic hardware simulator, and for this work it is configured with parameters to match the  $\mu$ Caspian hardware implementation with equivalent deterministic behavior. The  $\mu$ Caspian NEP core is the target SNN hardware accelerator. It is built into the embedded neuromorphic platform’s FPGA as presented in [23] and used for edge deployment and power analysis in this work.

$\mu$ Caspian is specifically designed to fit within the constrained resources of a Lattice iCE40UP5K FPGA. This FPGA is small, low-power, and inexpensive, making it well suited for low-power edge applications.  $\mu$ Caspian implements the IF neuron model with 8-bit thresholds, 16-bit charge state, and 4-bit axonal delay (0-15 cycles). Each synapse has an 8-bit signed weight value.  $\mu$ Caspian only supports axon delay, without support for synaptic delay. Spike routing is handled through target neuron lookup in the synapse routing tables.

This allows for very flexible routing within the SNN network, with support for self-connections, recurrent connections, and up to an all-to-all connection pattern. The max size of an SNN is limited by the memory used to store the SNN configurations, allowing up to 256 neurons and 4096 synapses, with any connection pattern.

The  $\mu$ Caspian NEP is an event-driven pipeline that mirrors the SNN model’s stages: communication, control, synapses, dendrites, neurons, and axons. In each cycle, dendrite charges move to neurons, which check for spikes, add any configured axon delays, and pass spikes through synapses back into dendrites. Because all stages run in parallel and only process events, the time and power per cycle vary with network activity, internal states, and inputs.

### C. Training

1) *EONS*: In this work, we train SNNs with the EONS approach. As an evolutionary algorithm, EONS begins with a population of candidate SNNs, denoted by

$$\mathcal{P}^{(g)} = \{S_1^{(g)}, S_2^{(g)}, \dots, S_N^{(g)}\},$$

where  $S_i^{(g)}$  represents the  $i$ -th SNN in generation  $g$  and  $N$  is the population size. Each SNN  $S_i^{(g)}$  is either randomly initialized or initialized using pre-created networks. These networks are evaluated using the fitness function  $F$  (described below), and each network is assigned a fitness score  $F(S_i^{(g)})$ . These fitness scores are used to inform the selection process, which is used to choose which networks will serve as parents for the next generation. In this work, we use *tournament selection*: a random subset of individuals,

$$\mathcal{T} = \{S_{j_1}^{(g)}, \dots, S_{j_k}^{(g)}\} \subseteq \mathcal{P}^{(g)},$$

compete in a tournament, and then the best individual in that subset is selected as the parent. Formally, if  $\mathcal{T}$  contains  $k$  individuals, then the parent chosen from  $\mathcal{T}$  is

$$\tilde{S}^{(g)} = \arg \max_{S \in \mathcal{T}} F(S).$$

This process allows for a nice tradeoff between preferentially selecting better performing networks while still allowing for the chance that lower performing individuals are also occasionally selected (i.e., in the case that the tournament is entirely made up of lower performing individuals).

Once parents are selected, reproduction operations are applied to generate the next generation  $\mathcal{P}^{(g+1)}$ . EONS uses three different reproduction operators:

- *Duplication (Cloning)*: A straightforward copy of a parent  $\tilde{S}^{(g)}$  into the next generation.
- *Crossover*: Given two parents,  $\tilde{S}_{p_1}^{(g)}$  and  $\tilde{S}_{p_2}^{(g)}$ , the offspring inherit both structural and parameter characteristics from each parent.
- *Mutation*: With some probability, a parent’s architecture or parameters are altered. Possible mutations include adding or removing neurons/synapses or perturbing parameters such as synaptic weights or neuronal thresholds.

TABLE I  
SELECTED EONS HYPERPARAMETERS WITHIN TENNLAB

Parameter	Value
starting_nodes	50
starting_edges	50
merge_rate	0
population_size	100
multi_edges	0
crossover_rate	0.5
mutation_rate	0.9
tournament_size_factor	0.1
tournament_best_net_factor	0.9
random_factor	0.05
num_mutations	4
node_mutations	{Threshold: 1.0}
edge_mutations	{Weight: 0.7, Delay: 0.3}
num_best	3
add_node_rate	0.5
delete_node_rate	0.25
add_edge_rate	0.75
delete_edge_rate	0.25
node_params_rate	2.5
edge_params_rate	2.5

EONS also uses elitism which guarantees that some number of the best performing individuals from  $\mathcal{P}^{(g)}$  are cloned into  $\mathcal{P}^{(g+1)}$ . Once the children have been created, they replace the parent population and are evaluated using the fitness function. From there, this process of selection, reproduction, and evaluation continues over the course of a pre-specified number of generations or a fixed amount of time. The final population after training is completed is denoted as  $\mathcal{P}^{(*)}$ .

Table I contains the values of the EONS hyperparameters used for all of the applications discussed in Section IV. The meaning of each of these parameters is outlined in the EONS documentation and prior work [14].

When training an SNN through EONS, our objective is to optimize both  $\Psi$  and  $f$  to maximize a fitness function  $F$ . This involves two key components:

- Evolving an architecture and the weights that enable  $\Psi$  to effectively encode salient historical features in the model's internal state  $\mathcal{H}$ , maintaining short-term memory of relevant information.
- Developing a function  $f$  that utilizes this short-term memory ( $\mathcal{H}$ ) and the current observation ( $\mathbf{x}_t$ ) to classify the input as either 0 or 1.

During training, the threshold  $\theta$  is set to 0, to encourage the SNN to only produce output spikes during class-1 encounters. This differs in evaluation, as described in Section III-D.

2) *Fitness Function Selection*: The choice of  $F$  plays a critical role in training. With EONS, the developer has the freedom to design a fitness function without the restrictions of differentiability imposed by training algorithms based on gradient descent. As such, metrics based directly on the discrete binary performance, such as classification accuracy, precision, recall, or F1-score, can be directly incorporated into the fitness function.

For most applications, we have found that the Matthews Correlation Coefficient (MCC) performs well as a fitness function. The MCC is a metric that measures the correlation

between the true and predicted binary classes and considers all four components of the binary confusion matrix. It is robust to unbalanced datasets and is regarded as one of the most informative measures of a binary classifier's performance [24]–[26]. The MCC is defined as

$$\text{MCC} = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}}$$

and ranges from  $-1$  (perfect inverse correlation) to  $+1$  (perfect correlation), with 0 indicating no correlation.

Metrics that are sensitive to imbalanced data, such as the Area Under the Receiver Operating Characteristic Curve (AUC), were found to be poor choices for fitness functions. SNNs trained for AUC tended to converge to a local minimum of predicting the majority class. Data balancing procedures may mitigate this issue; however, one could simply select metrics like the MCC, that are inherently robust to imbalance.

#### D. Evaluation

One of the benefits of EONS is that repeated training procedures can produce a population of diverse candidate networks, many of which may be suitable for deployment. From our testing, we have found that it is best to train more than one population of networks for a given application. Depending on the application and computing resources available, anywhere from 100 to 1000 populations are trained. The resulting SNNs from these populations form their own trained population, denoted by  $\mathcal{P}^{(*)} = \{S_1^{(*)}, \dots, S_N^{(*)}\}$ . Each SNN in  $\mathcal{P}^{(*)}$  is evaluated on the training set using a variety of standard metrics (e.g., accuracy, precision, recall, MCC). All possible values of the threshold,  $\theta$ , are evaluated, and the one resulting in the maximum MCC is then used for sorting the SNNs in  $\mathcal{P}^{(*)}$ . The top 10% of these are then considered for network ensembling.

#### E. Network Ensembling

We have observed that the top SNNs often have different strengths and weaknesses. We employ *network ensembling* through simple voting procedures to leverage each model's strengths. Specifically, once we identify the top networks based on their training-set MCC scores, we evaluate all pairs and trios drawn from these top performers. The ensemble with the highest training set score is then evaluated on the test set.

For any given input, the ensemble's prediction is obtained via either:

- *Any Vote*: The ensemble outputs a class-1 prediction if *any* of the networks predict class 1.
- *Majority Vote*: The ensemble outputs class 1 if at least two out of the three networks predict class 1.

These voting mechanisms allow us to capitalize on complementary decision boundaries learned by different networks, often resulting in improved overall performance. While more sophisticated ensembling approaches are available, simple voting procedures have the benefit of maintaining interpretability and minimizing computational complexity. They also provide developers the freedom to operate the ensemble members either serially or in parallel.

## IV. APPLICATIONS

### A. Source Detection in Gamma-ray Spectral Data

1) *Background*: In this application, we train SNNs to detect radioactive sources in gamma-ray spectral time series data. We define a *source* as any radioactive object or material that is not part of the natural radioactive *background*. The intensity and energy distribution (spectrum) of gamma radiation measured by a sensor can vary significantly due to differences in the concentrations of naturally occurring radioactive nuclides. Consequently, the background signal is highly variable in both magnitude and shape, which makes identifying anomalous signals (i.e., potential sources) more challenging.

Our study uses a spectroscopic detector, specifically a NaI(Tl) crystal measuring 3"×3" in volume, which records the energy deposited by incoming gamma-rays. Each source of interest exhibits a unique set of gamma-ray energies, and these can be exploited for both detection and identification. However, the intensity of the gamma-ray signature decreases according to an inverse square law ( $1/r^2$ ) with respect to the distance  $r$  between source and detector, and any intervening shielding can further attenuate the signal.

This combination of factors introduces significant variability into the measured spectra, resulting in a low SNR. In prior work, detection algorithms built on ANNs have demonstrated strong performance in this application [27], [28]; however, their energy consumption and latency can limit their viability for long-term edge-based operation.

2) *Dataset*: Building on the experimentally validated framework used for generating synthetic datasets in prior work [29], [30], we created a new synthetic dataset featuring additional sources and greater background variation. This dataset simulates a 3"×3" NaI(Tl) detector moving at walking speed through a virtual model of the 7000-area at ORNL. We simulate 20 unique source/shielding combinations, each at 8 distinct SNR levels. Here,  $\text{SNR} = \frac{S}{\sqrt{S+B}}$ , where  $S$  is the total source counts and  $B$  is the total background counts measured during an encounter. To generate individual runs, we sample from the background distributions and simulate the detector passing each source at various speeds and standoff distances.

We also collected over 8 hours of real-world dynamic background radiation spectra by driving a detector around the Knoxville, Tennessee region and recording data. Half of these data were included in training and half in testing. Synthetic sources were also injected over random sub-ranges of the measured background runs. Table II below provides a breakdown of the datasets.

3) *Data Preprocessing*: Each run in the dataset consists of a time series of gamma-ray spectra, with each spectrum represented by 32 energy bins. The integer value in each bin corresponds to the number of gamma rays detected in that bin's energy range at a given time step. Prior to spike encoding, each spectrum is normalized to an integral sum of one. We then generate spike trains using *rate* encoding, as described in Section III. This method was chosen because it performed slightly better than *spikes* encoding for this application [15].

TABLE II  
DATASET PARAMETERS FOR THE GAMMA-RAY SOURCE DETECTION APPLICATION

Parameter	Training	Test
<i>Spectral Integration</i>		
Integration Time	0.5 s	–
Integration Stride	0.5 s	–
Calibration Range	1 keV– 3001 keV	–
Number of Energy Bins	32	–
Binning Structure	$\sqrt{E}$	–
Background Variability	80% about nominal [29]	–
<i>Sources</i>		
SNM Sources	25 kg HEU, 8 kg WGPu	–
Point Sources	$^{241}\text{Am}$ , $^{133}\text{Ba}$ , $^{57}\text{Co}$ , $^{60}\text{Co}$ , $^{137}\text{Cs}$ , $^{177}\text{Lu}$	–
SNM Shielding	bare, steel, Cu, Pb	–
Point Source Shielding	bare, steel	–
Source SNR	8 steps up to 16	8 steps up to 18
<i>Speed &amp; Offset</i>		
Relative Speed (Synthetic)	0.5–1.5 m/s	0.5–2.2 m/s
Detector/Source Offset (Synthetic)	2 m, 4 m	–
Relative Speed (Injected)	2.2–13.4 m/s	–
<i>Dataset Size</i>		
Background-only	240 runs	–
Containing Source	3520 runs	–

4) *Training*: We train 1000 populations of 100 networks using the procedure described in Section III. This application demands a very low FPR, consistent with standards such as ANSI N42.53 [31] and ANSI N42.35 [32], which specify maximum False Alarm Rates (FARs) of  $1\text{ hr}^{-1}$  and  $0.5\text{ hr}^{-1}$ , respectively. The International Atomic Energy Agency further recommends a maximum FAR of  $1/12\text{ hr}^{-1}$  [33].

To meet these requirements, we define a metric, denoted  $\text{TPR}_0$ , which is the true positive rate of an SNN at a FAR of  $0\text{ hr}^{-1}$ . Our fitness function is then:

$$F = \text{F1} + \text{TPR}_0^2$$

where  $\text{F1}$  is the harmonic mean of precision and recall. This metric is inspired by that used in [16]. After each training epoch, Receiver Operating Characteristic (ROC) curves are generated for each network, from which  $\text{TPR}_0$  is derived.

Initially, we attempted using  $F = \text{TPR}_0$  alone, but during early training stages, most networks could not detect any sources at a FAR of  $0\text{ hr}^{-1}$ . As a result,  $\text{TPR}_0$  was zero for nearly all networks, causing the training process to stagnate. Incorporating  $\text{F1}$  into the objective enables the networks to reach a non-zero  $\text{TPR}_0$  and thus effectively optimize both the overall detection performance and the low FAR requirement.

Due to the large size of our training set, each epoch uses a training batch that is a random 1% subset of the whole. We also implement a new feature, referred to as *SNR scaling*, where we gradually increase the difficulty of this training batch for each epoch. We do this by biasing samples toward high SNR early in training and sampling uniformly in later epochs.

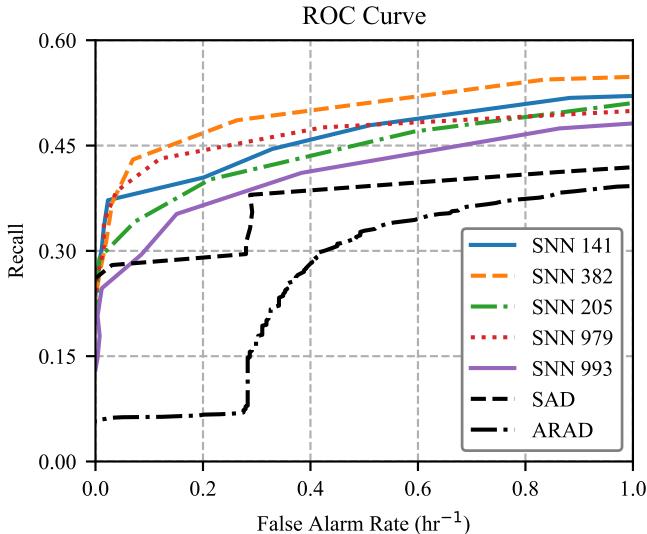


Fig. 1. Testing dataset ROC curves for the top 5 SNNs and reference ARAD and SAD algorithms.

After the 1000 populations were trained and evaluated on the test set, the top 10 networks were selected for ensembling evaluation. The optimal thresholds for these networks were set to yield an ensemble FAR  $\leq 1\text{hr}^{-1}$ .

5) *Comparisons:* We compare our results to two well-established algorithms, the Spectral Anomaly Detector (SAD) [34], [35] and the Autoencoder Radiation Anomaly Detector (ARAD) [28]. SAD is a simple but effective source detection algorithm that learns a low-dimensionality linear representation of background spectra (32 energy bins) using Principal Component Analysis (PCA), using spectral reconstruction error to detect non-background sources. ARAD operates in a similar manner to SAD, but utilizes a very small ( $\sim 3\text{k}$  parameters) deep convolutional autoencoder to learn a nonlinear representation of background radiation spectra (128 energy bins). Both the ARAD and SAD algorithms perform binary classification under the *Independent Sample Classification* paradigm outlined in Section I. Operationally, SAD provides the best alternative to SNNs in a low-power setting, as it can easily be implemented on a MCU, as described in Section V-B. ARAD has been shown to outperform SAD in highly dynamic background source detection scenarios when utilizing large NaI(Tl) detectors [28].

### B. Seizure Detection in EEG

We also evaluate our approach is seizure detection in Electroencephalogram (EEG) data. We use the CHB-MIT Scalp EEG database [36]–[38], a publicly available dataset of pediatric EEG recordings collected at Children’s Hospital Boston. It comprises 686 continuous EEG records from 23 pediatric subjects with intractable seizures, divided into 24 cases. Each case contains between 9 and 42 continuous .edf

Dataset	Patients	Seizure Records	Total
Training Set	chb12, chb15, chb04, chb01, chb05, chb03, chb02, chb18, chb13	100	368
Validation Set	chb11, chb06, chb10	18	80
Test Set	chb21, chb20, chb22, chb19, chb14, chb24, chb16, chb17, chb08, chb09, chb07, chb23	71	294

TABLE III  
DATASET SPLITS ACROSS PATIENTS AND RECORDS.

records. The dataset includes 198 annotated seizures and details the onset and duration of each event.

These recordings include high-resolution signals sampled at 256 Hz, with most containing 23 EEG channels following the International 10-20 electrode placement system. The data is continuously recorded [39] for durations of 1, 2, and 4-hours.

Unlike our in-depth analysis of gamma-ray spectral data, the intent behind the examination of this dataset is to evaluate the performance of our general methodology on a new application without implementing any domain-specific optimizations.

1) *Data Preprocessing:* To ensure consistency across records, data from 18 sensors were used as the SNN input. Records containing multiple seizures were divided into separate runs, each featuring a single seizure. This resulted in a dataset composed of 189 seizures across 742 records, with 2 being excluded due to missing or inconsistent annotations.

The records were trimmed to a maximum duration of 400 seconds, fully encapsulating the seizure events, keeping an amount of padding at the beginning and end of the seizure. This padding was chosen randomly between 8 and 20 s to introduce variability in the start times of the seizure events. This trimming procedure was performed to fully capture the temporal dynamics of seizure events while minimizing record length for computational efficiency. Records with a duration shorter than 400 seconds were unmodified.

The dataset was then divided across patients and records into training (50%), validation (10%), and testing (40%) subsets. The test set is fully comprised of patients not seen in training or validation. Table III presents the final dataset split statistics.

2) *Training:* Training follows the general procedure in Section III. To account for the high sampling rate, we apply a 20 s rolling sum over the spike output during testing (evaluation only), with the detection threshold applied to this sum. This window length was chosen empirically to encompass seizure duration. The rolling sum is omitted during training to better encourage the network to suppress false positives.

3) *Comparisons:* Several alternative approaches have been evaluated on the CHB-MIT Scalp EEG dataset, often incorporating specialized preprocessing or model architectures that yield a range of performance outcomes. For instance, [40] employs a fully connected ANN feeding into an Long Short-term Memory (LSTM), achieving a sensitivity of 97.746 and a FPR of 0.2373 across 21 patients. Similarly, [41] apply frequency-based techniques to further refine feature extraction, achieving a reported sensitivity of 93.

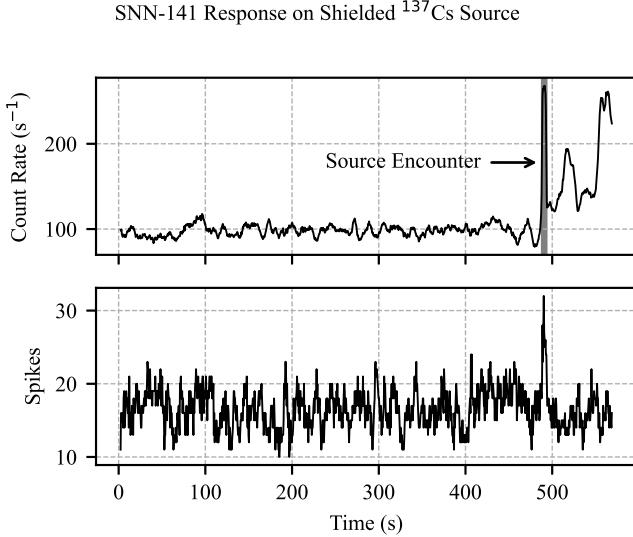


Fig. 2. Spike response of *SNN 141* on a time series from the testing dataset, with a synthetic shielded  $^{137}\text{Cs}$  source injected on measured background.

## V. RESULTS AND DISCUSSION

### A. Source Detection in Gamma-ray Spectral Data

Fig.1 shows the ROC curves generated on the full testing dataset, for the top 5 networks, as defined by their TPR at a FAR of  $1\text{hr}^{-1}$  on the training set. The ROC curve is shown for FARs from 0 to  $2\text{hr}^{-1}$ , displaying the full range of relevance based on United States and international recommendations [31], [32], [42]. The ROC curves for the reference SAD and ARAD algorithms are also shown.

The top performing SNN was *SNN 141*, with  $\text{TPR}_1 = 0.498$ . *SNN 141* comprises 49 neurons and 66 synapses, with a highly recurrent architecture. Fig.2 shows the spike response of *SNN 141* on a run from the testing set containing a synthetic shielded  $^{137}\text{Cs}$  source injected over real-world background. On the test set,  $\text{TPR}_1 = 0.518$  for *SNN 141*.

All unique ensembles of the top-100 networks from training were evaluated and sorted according to their  $\text{TPR}_1$  scores on the training set. The top-780 ensembles were all trio ensembles, with the highest performing one reaching  $\text{TPR}_1 = 0.671$ , a 29.5% improvement. The highest performing pair ensemble reached  $\text{TPR}_1 = 0.647$ , a 24.9% improvement over *SNN 141*. All ensembles that outperformed *SNN 141* used the *any* voting method. Fig.3 shows the performance of the top ensemble compared against its individual constituents, *SNN 135*, *SNN 33*, and *SNN 75*, on the non-shielded sources in the test set.

1) *Discussion:* As described in [16], many of the sources are simulated at very low SNRs, thus very high performance on this dataset may not be achievable. The results should be interpreted relative to the reference algorithm performance, which have been established in the community.

As shown in Fig. 1, the top five SNNs outperformed both SAD and ARAD, which work well on larger detectors but are less robust for smaller ones. Smaller detectors yield fewer

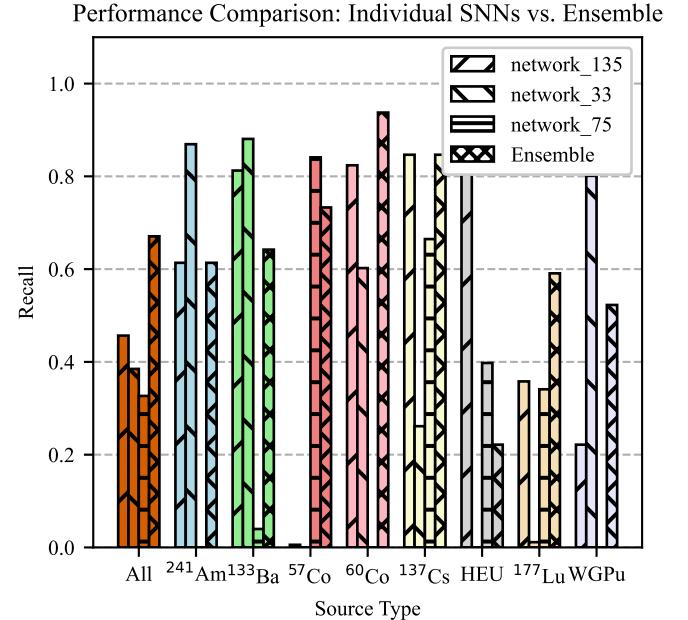


Fig. 3. Recall for the top ensemble and its individual constituents for all non-shielded sources and all sources together (All).

counts, increasing relative noise from Poisson statistics. We hypothesize that the SNNs' ability to model both spectral and temporal patterns improves noise robustness and temporal pattern recognition, aligning with [16], which highlighted the benefits of short-term memory in this domain.

Fig. 2 illustrates the spiking behavior of the top-performing SNN under a variety of conditions: from 0 to about 480 s, the background remains relatively steady with no source present; around 480–490 s, a shielded  $^{137}\text{Cs}$  source is introduced; after 490 s, only background is observed, but it becomes highly dynamic, likely due to the detector passing buildings. As shown, the SNN exhibits low output neuron spiking activity during both steady and dynamic background conditions, while sharply increasing with source presence.

Finally, ensembling either two or three of the top SNNs boosted performance by more than 15% over single-network results. Fig.3 demonstrates how *any* vote ensembling allows the ensemble to leverage the collective strengths of each of the constituent SNNs, with different networks performing better than others on different source types.

In the case of  $\mu$ Caspian where each network in the ensemble is run sequentially, ensembling would increase latency and energy consumption roughly proportional to the number of networks in the ensemble. This trade-off may be viable, however, for applications that can tolerate additional overhead.

### B. Power Analysis

A power analysis was conducted to compare our SNN implementation with the SAD algorithm on a custom embedded platform that integrates an STM32U5 MCU and a Lattice iCE40 UltraPlus FPGA. The  $\mu$ Caspian SNN core resides

in the FPGA, while the MCU handles spike encoding and decoding. This custom hardware platform was designed with the algorithmic methods described herein, and is described in greater detail in prior work [23].

The SAD algorithm runs entirely on the MCU using the ARM CMSIS-DSP library with single-precision floating-point arithmetic. Power measurements were performed with a Keysight N6705C analyzer, bypassing on-board regulators to power the 1.2V (FPGA core) and 3.3V (I/O and MCU) rails separately. When idle, the MCU consumes about 6mW, and the FPGA clock can be disabled to 300 $\mu$ W. During inference, SAD executes in roughly 100 $\mu$ s (plus idle-transition overhead) at 160MHz, requiring an estimated 27  $\mu$ J per inference.

In contrast, the SNN consumes about 2mW at 24MHz for 20.2ms during inference, or 40.4 $\mu$ J, plus an additional 32mW (646.4 $\mu$ J) while transferring spikes between the MCU and FPGA. This overhead occurs because software-based encoding keeps the MCU from entering low-power states.

*1) Discussion:* These results demonstrate that power management is a complex co-design issue: both the MCU and FPGA remain in high-power states during inference. Future improvements involve reducing the MCU clock, refining the communication protocol, and leveraging DMA or lower-power communication to reduce data-movement overhead. Finally, while this system currently supports hot-swapping of SNNs for serial ensemble member operation, future work will support parallel operation of multiple SNNs with shared inputs.

#### C. Seizure Detection in EEG

The testing set results are presented in Table IV. The highest performing single model achieved an MCC score of 0.47 and was comprised of 18 neurons connected by 23 synapses with high recurrence. Ensembling proved to be highly effective for enhancing network performance, with trio ensembles achieving the highest scores. Performance changes in MCC scores were observed as follows: +2.13% for the pair ensemble, -34.04% for the trio any-vote ensemble, and +51.06% for the trio majority-vote ensemble.

The pair ensemble demonstrated only a modest gain in performance. The any-vote method for the trio ensemble suffered a decline in MCC and an increase in FPR. In contrast, the majority-vote ensemble yielded significantly higher performance, achieving the best overall performance with a TPR of 0.95 and an FPR of 0.16.

The majority-vote trio ensemble results are competitive with prior work on the CHB-MIT EEG dataset. For instance, [40] reported a TPR of 0.97 and an FPR of 0.24, while [41] achieved a TPR of 0.93 and an FPR of 0.06. However, both approaches relied on extensive preprocessing and artifact extraction, whereas our method achieved comparable performance using raw data.

These results demonstrate the generalizability of our methodology with minimal application-specific modifications. For maximum performance on specific applications, one should consider domain-specific design choices in fitness function and data preprocessing/encoding. This is standard practice for any

Model	Ensemble Method	TP	TN	FP	FN	TPR	FPR	MCC
Individual	–	27	212	8	47	0.36	0.04	0.47
Pair	vote	25	216	4	49	0.33	0.02	.48
Trio	any vote	32	167	53	42	0.43	0.24	0.31
Trio	majority vote	70	184	36	4	0.95	0.16	<b>0.71</b>

TABLE IV  
EEG TESTING SET PERFORMANCE METRICS FOR THE TOP INDIVIDUAL SNN, TOP PAIR ENSEMBLE, AND TOP TRIO ENSEMBLE.

AI application and can be automated using hyperparameter optimization tools.

## VI. CONCLUSION

In this work, we proposed a general-framework for training and evaluating SNNs for binary classification in multivariate time series. We performed a thorough validation of this framework, with application-specific optimizations, on a challenging dataset involving the detection of low SNR radioactive sources, comparing against alternative deep learning and PCA-based approaches. We then tested the generalizability of the approach without adapting it for specific applications, by evaluating it on seizure detection in EEG waveforms. Based on our results, we make the following conclusions:

- 1) The proposed framework is an effective solution for binary classification tasks in multivariate time series, with the resulting networks being directly deployable on the  $\mu$ Caspian neuromorphic hardware platform.
- 2) With application-specific optimizations on data pre-processing, spike encoder selection, and fitness function selection, the framework can yield SNNs with performance competitive with or exceeding alternative low-resource algorithms. This is evident in the results on the radiation dataset, where the SNN outperformed both the PCA and deep learning approach.
- 3) EONS offers the benefit of evolving SNNs with diverse capabilities, allowing one to leverage simple voting ensembling across two or three SNNs to yield significant performance gains.
- 4) The low-power  $\mu$ Caspian neuromorphic computing platform is a good choice for this domain, performing inference on  $\sim$ 2 mW of power.

## ACKNOWLEDGMENT

This research used resources of the Experimental Computing Laboratory (ExCL) at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725.

Portions of this research were conducted with high performance computational resources provided by Louisiana State University (<http://www.hpc.lsu.edu>) and also the Louisiana Optical Network Infrastructure (<http://www.loni.org>).

## REFERENCES

- [1] E. Strubell, A. Ganesh, and A. McCallum, “Energy and policy considerations for modern deep learning research,” in *Proceedings of the AAAI conference on artificial intelligence*, vol. 34, no. 09, 2020, pp. 13 693–13 696.

- [2] P. Villalobos, A. Ho, J. Sevilla, T. Besiroglu, L. Heim, and M. Hobbehahn, "Position: Will we run out of data? limits of LLM scaling based on human-generated data," in *Forty-first International Conference on Machine Learning*, 2024. [Online]. Available: <https://openreview.net/forum?id=ViZcgDQjyG>
- [3] K. Valmeeakam, A. Olmo, S. Sreedharan, and S. Kambhampati, "Large language models still can't plan (a benchmark for llms on planning and reasoning about change)," in *NeurIPS 2022 Foundation Models for Decision Making Workshop*, 2022.
- [4] I. Mirzadeh, K. Alizadeh, H. Shahrokhi, O. Tuzel, S. Bengio, and M. Farajtabar, "Gsm-symbolic: Understanding the limitations of mathematical reasoning in large language models," *arXiv preprint arXiv:2410.05229*, 2024, unpublished.
- [5] B. Jiang, Y. Xie, Z. Hao, X. Wang, T. Mallick, W. J. Su, C. J. Taylor, and D. Roth, "A peek into token bias: Large language models are not yet genuine reasoners," *arXiv preprint arXiv:2406.11050*, 2024, in press.
- [6] K. Roy, A. Jaiswal, and P. Panda, "Towards spike-based machine intelligence with neuromorphic computing," *Nature*, vol. 575, no. 7784, pp. 607–617, 2019.
- [7] C. D. Schuman, S. R. Kulkarni, M. Parsa, J. P. Mitchell, P. Date, and B. Kay, "Opportunities for neuromorphic computing algorithms and applications," *Nature Computational Science*, vol. 2, no. 1, pp. 10–19, 2022.
- [8] A. Calimera, E. Macii, and M. Poncino, "The human brain project and neuromorphic computing," *Functional neurology*, vol. 28, no. 3, p. 191, 2013.
- [9] C. D. Schuman, T. E. Potok, R. M. Patton, J. D. Birdwell, M. E. Dean, G. S. Rose, and J. S. Plank, "A survey of neuromorphic computing and neural networks in hardware," *arXiv preprint arXiv:1705.06963*, 2017, unpublished.
- [10] O. Krestinskaya, A. P. James, and L. O. Chua, "Neuromemristive circuits for edge computing: A review," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 1, pp. 4–23, 2019.
- [11] J. P. Mitchell, C. D. Schuman, R. M. Patton, and T. E. Potok, "Caspian: A neuromorphic development platform," in *Proceedings of the Neuro-Inspired Computational Elements Workshop*, ser. NICE '20. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: <https://doi.org/10.1145/3381755.3381764>
- [12] E. O. Neftci, H. Mostafa, and F. Zenke, "Surrogate gradient learning in spiking neural networks: Bringing the power of gradient-based optimization to spiking neural networks," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 51–63, 2019.
- [13] T. C. Wunderlich and C. Pehle, "Event-based backpropagation can compute exact gradients for spiking neural networks," *Scientific Reports*, vol. 11, no. 1, p. 12829, 2021.
- [14] C. D. Schuman, J. P. Mitchell, R. M. Patton, T. E. Potok, and J. S. Plank, "Evolutionary optimization for neuromorphic systems," in *Proceedings of the 2020 Annual Neuro-Inspired Computational Elements Workshop*, 2020, pp. 1–9.
- [15] J. Ghawaly, A. Young, D. Archer, N. Prins, B. Witherspoon, and C. Schuman, "A neuromorphic algorithm for radiation anomaly detection," in *Proceedings of the International Conference on Neuromorphic Systems 2022*, 2022, pp. 1–6.
- [16] J. Ghawaly, A. Young, A. Nicholson, B. Witherspoon, N. Prins, M. Swinney, C. Celik, C. Schuman, and K. Patel, "Performance optimization study of the neuromorphic radiation anomaly detector," in *Proceedings of the 2023 International Conference on Neuromorphic Systems*, 2023, pp. 1–7.
- [17] P. S. Maciąg, M. Kryszkiewicz, R. Bembeniak, J. L. Lobo, and J. Del Ser, "Unsupervised anomaly detection in stream data with online evolving spiking neural networks," *Neural Networks*, vol. 139, pp. 118–139, 2021.
- [18] D. Bäßler, T. Kortus, and G. Gühring, "Unsupervised anomaly detection in multivariate time series with online evolving spiking neural networks," *Machine Learning*, vol. 111, no. 4, pp. 1377–1408, 2022.
- [19] Y. Cherdo, B. Miramond, and A. Pegatoquet, "Time series prediction and anomaly detection with recurrent spiking neural networks," in *2023 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2023, pp. 1–10.
- [20] R. Gaurav, T. C. Stewart, and Y. Yi, "Reservoir based spiking models for univariate time series classification," *Frontiers in Computational Neuroscience*, vol. 17, p. 1148284, 2023.
- [21] G. Neculae, O. Rhodes, and G. Brown, "Ensembles of spiking neural networks," *arXiv preprint arXiv:2010.14619*, 2020.
- [22] J. P. Mitchell, C. D. Schuman, and T. E. Potok, "A small, low cost event-driven architecture for spiking neural networks on fpgas," in *International Conference on Neuromorphic Systems 2020*, ser. ICONS 2020. New York, NY, USA: Association for Computing Machinery, 2020. [Online]. Available: <https://doi.org/10.1145/3407197.3407216>
- [23] B. Witherspoon and A. Young, "Event-driven sensing and embedded neuromorphic platforms for gamma radiation monitoring," in *Proceedings of the Great Lakes Symposium on VLSI 2024*, 2024, pp. 779–784.
- [24] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, pp. 1–13, 2020.
- [25] D. Chicco, N. Tötsch, and G. Jurman, "The matthews correlation coefficient (mcc) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation," *BioData mining*, vol. 14, pp. 1–22, 2021.
- [26] S. Boughorbel, F. Jaray, and M. El-Anbari, "Optimal classifier for imbalanced data using matthews correlation coefficient metric," *PloS one*, vol. 12, no. 6, p. e0177678, 2017.
- [27] C. M. Anderson-Cook, D. Archer, M. S. Bandstra, J. C. Curtis, J. M. Ghawaly, T. H. Joshi, K. L. Myers, A. D. Nicholson, and B. J. Quiter, "Radiation detection data competition report," Los Alamos National Laboratory (LANL), Los Alamos, NM (United States), Tech. Rep., 2021.
- [28] J. M. Ghawaly Jr, A. D. Nicholson, D. E. Archer, M. J. Willis, I. Garishvili, B. Longmire, A. J. Rowe, I. R. Stewart, and M. T. Cook, "Characterization of the autoencoder radiation anomaly detection (arad) model," *Engineering Applications of Artificial Intelligence*, vol. 111, p. 104761, 2022.
- [29] J. M. Ghawaly, A. D. Nicholson, D. E. Peplow, C. M. Anderson-Cook, K. L. Myers, D. E. Archer, M. J. Willis, and B. J. Quiter, "Data for training and testing radiation detection algorithms in an urban environment," *Scientific data*, vol. 7, no. 1, pp. 1–6, 2020.
- [30] A. D. Nicholson, D. E. Peplow, J. M. Ghawaly, M. J. Willis, and D. E. Archer, "Generation of synthetic data for a radiation detection algorithm competition," *IEEE Transactions on Nuclear Science*, vol. 67, no. 8, pp. 1968–1975, 2020.
- [31] "American national standard performance criteria for backpack-based radiation-detection systems used for homeland security," *ANSI N42.53-2013*, pp. 1–48, 2013.
- [32] "American national standard for evaluation and performance of radiation detection portal monitors for use in homeland security," *ANSI N42.35-2016 (Revision of ANSI N42.35-2006)*, pp. 1–70, 2016.
- [33] "Detection of radioactive materials at borders," International Atomic Energy Agency, Tech. Rep., 2002.
- [34] K. Miller and A. Dubrawski, "Gamma-ray source detection with small sensors," *IEEE Transactions on Nuclear Science*, vol. 65, no. 4, pp. 1047–1058, 2018.
- [35] P. Tandon, P. Huggins, R. Maclachlan, A. Dubrawski, K. Nelson, and S. Labov, "Detection of radioactive sources in urban scenes using bayesian aggregation of data from mobile spectrometers," *Information Systems*, vol. 57, pp. 195–206, 2016.
- [36] J. Guttag, "CHB-MIT Scalp EEG Database (version 1.0.0)," PhysioNet, 2010. [Online]. Available: <https://doi.org/10.13026/C2K01R>
- [37] A. Shoeb, "Application of machine learning to epileptic seizure onset detection and treatment." Ph.D. dissertation, Massachusetts Institute of Technology, September 2009.
- [38] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. C. Ivanov, R. Mark, and H. E. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000, [Online].
- [39] S. Wong, A. Simmons, J. Rivera-Villicana, S. Barnett, S. Sivathamboo, P. Perucca, Z. Ge, P. Kwan, L. Kuhlmann, R. Vasa, K. Mouzakis, and T. J. O'Brien, "EEG datasets for seizure detection and prediction—A review," *Epilepsia Open*, vol. 8, no. 2, pp. 252–267, 2023.
- [40] A. Bhattacharya, "Automatic seizure prediction using cnn and lstm," *arXiv preprint arXiv:2211.02679*, 2022.
- [41] P. Fergus, D. Hignett, A. Hussain, D. Al-Jumeily, and K. Abdel-Aziz, "Automatic epileptic seizure detection using scalp eeg and advanced artificial intelligence techniques," *BioMed research international*, vol. 2015, no. 1, p. 986736, 2015.
- [42] "American national standard performance criteria for handheld instruments for the detection and identification of radionuclides," *ANSI N42.34-2015 (Revision of ANSI N42.34-2006)*, pp. 1–60, 2016.